CHAPTER 3

Molecular Docking
3.1 Introduction to molecular docking

Molecular docking technique is the famous in structural bioinformatics to solve the problems in protein and ligand interaction studies. A molecule is characterized by a pair \((A; B)\), in which \(A\) represents a collection of atoms, and \(B\) represents a collection of bonds between pairs of atoms. Information used for kinematic and energy computations is associated with each of the atoms and bonds. Each atom carries standard information, such as its van der Waals radius. Three pieces of information are associated with each bond: (i) the bond length is the distance between atom centers; (ii) the bond angle, is the angle between two consecutive bonds; (iii) whether the bond is rotatable or not (Fig. 3.1). Since bond lengths and angles do not affect significantly the shape of a molecule, it is common practice to consider them fixed. Thus the degrees of freedom of the molecule arise from the rotatable bonds. The three dimensional embedding of a molecule defined when we assign values to its rotatable bonds is called the conformation of the molecule. Ligands typically have 3 to 15 rotatable bonds, while receptors have 1,000 to 2,000 rotatable bonds. The dimension of the combined searched space makes the docking problem computationally intractable.

![Fig 3.1- A drug molecule. Spheres represent atoms and bonds connecting them are represented by sticks. Curved arrows represent the rotatable degrees of freedom around bonds.](image)

One key aspect of molecular modeling is calculating the energy of conformations and interactions. This energy can be calculated with a wide range of methods are ranging from quantum mechanics to purely empirical
energy functions. The accuracy of these functions is usually proportional to its computational expense and choosing the correct energy calculation method is highly dependent on the application. Computation times for different methods can range from a few milliseconds on a workstation to several days on a massively parallel supercomputer.

In the context of docking, energy evaluations are usually carried out with the help of a scoring function and developing these is a major challenge facing structure based drug design (Vieth et al., 1998). Scoring functions are a critical part of the structure based drug design process. No matter how efficient and accurate the geometric modeling of the binding process is, without good scoring functions it is impossible to obtain correct solutions. The two main characteristics of a good scoring function are selectivity and efficiency. Selectivity enables the function to distinguish between correctly and incorrectly docked structures and efficiency enables the docking program to run in a reasonable amount of time.

A large number of current scoring functions are based on forcefields that were initially designed to simulate the function of proteins (Cornell et al., 1995 and MacKerell et al., 1998). A force field is an empirical fit to the potential energy surface in which the protein exists and is obtained by establishing a model with a combination of bonded terms (bond distances, bond angles, torsional angles, etc.) and non-bonded terms (van der Waals and electrostatic). The relative contributions of these terms are adjusted for the different types of atoms in the simulated molecule by adjusting a series of parameters. Some scoring functions used in molecular docking have been adapted to include terms such as solvation and entropy (Morris et al., 1998). A separate approach is to use pure empirical scoring functions that are derived using multivariate regression methods of experimental data. In the market there are various docking softwares available among them Insight-II, MOE, GOLD and sybyl (Tripos) but most some of the softwares are academic free softwares like DOCK, Autodock, FlexX. From the literature we found Autodock is widely using software by academic institutions.

### 3.2 Autodock

The program AutoDock was developed to provide an automated procedure for predicting the interaction of ligands with biomacromolecular targets. The
motivation for development of Autodock software arises from problems in the design of bioactive compounds, and in particular the field of computer-aided drug design (CADD). Progress in biomolecular x-ray and NMR crystallography continues to provide a number of important protein and nucleic acid structures. These structures could be targets for bioactive agents in the control of animal and plant diseases, or simply key to understanding of a fundamental aspect of biology. The precise interaction of such agents or candidate molecules is important in the development process. Indeed, AutoDock can be a valuable tool in the x-ray structure determination process itself: given the electron density for a ligand, AutoDock can help to narrow the conformational possibilities and help identify a good structure. In any docking scheme two conflicting requirements must be balanced: the desire for a robust and accurate procedure, and the desire to keep the computational demands at a reasonable level. The ideal procedure would find the global minimum in the interaction energy between the substrate and the target protein, exploring all available degrees of freedom (DOF) for the system. However, it must also run on a laboratory workstation within an amount of time comparable to other computations that a structural researcher may undertake, such as a crystallographic refinement. In order to meet these demands a number of docking techniques simplify the docking procedure. Still one of the most common techniques in use today is manually-assisted docking. Here, the internal and orientational degrees of freedom in the substrate are under interactive control. While the energy evaluation for such techniques can be sophisticated, the global exploration of configuration space is limited. At the other end of the spectrum are automated methods such as exhaustive search and distance geometry. These methods can explore configuration space, but at the cost of a much simplified model for the energetic evaluation.

The original procedure developed for AutoDock used a Monte Carlo (MC) simulated annealing (SA) technique for configuration exploration with a rapid energy evaluation using grid-based molecular affinity potentials. It thus combined the advantages of exploring a large search space and a robust energy evaluation. This has proven to be a powerful approach to the problem of docking a flexible substrate into the binding site of a static protein. Input to
the procedure is minimal. The researcher specifies a rectangular volume around the protein, the rotatable bonds for the substrate, and an arbitrary or random starting configuration, and the procedure produces a relatively unbiased docking.

Rapid energy evaluation is achieved by precalculating atomic affinity potentials for each atom type in the substrate molecule in the manner described by Goodford (Goodford., 1985). In the AutoGrid procedure, the protein is embedded in a three-dimensional grid and a probe atom is placed at each grid point. The energy of interaction of this single atom with the protein is assigned to the grid point. An affinity grid is calculated for each type of atom in the substrate, typically carbon, oxygen, nitrogen and hydrogen, as well as a grid of electrostatic potential, either using a point charge of +1 as the probe, or using a Poisson-Boltzmann finite difference method, such as DELPHI 5, 6.

The energetic of a particular substrate configuration is then found by tri-linear interpolation of affinity values of the eight grid points surrounding each of the atoms in the substrate. The electrostatic interaction is evaluated similarly, by interpolating the values of the electrostatic potential and multiplying by the charge on the atom (the electrostatic term is evaluated separately to allow finer control of the substrate atomic charges). The time to perform an energy calculation using the grids is proportional only to the number of atoms in the substrate, and is independent of the number of atoms in the protein. The docking simulation is carried out using one of a number of possible search methods.

3.3 Free energy calculation with Autodock

The original AutoDock supported only one search method, although version 3.0 now has several. The original search algorithm was the Metropolis method, also known as Monte Carlo simulated annealing. With the protein static throughout the simulation, the lead molecule performs a random walk in the space around the protein. At each step in the simulation, a small random displacement is applied to each of the degrees of freedom of the substrate: translation of its center of gravity; orientation; and rotation around each of its flexible internal dihedral angles. This displacement results in a new configuration, whose energy is evaluated using the grid interpolation procedure described above. This new energy is compared to the energy of
the preceding step. If the new energy is lower, the new configuration is immediately accepted. If the new energy is higher, then the configuration accepted or rejected based upon a probability expression dependent on a user defined temperature, $T$. The probability of acceptance is given by:

$$P(\Delta E) = e^{-\frac{\Delta E}{k_BT}}$$

Where $\Delta E$ is the difference in energy from the previous step, and $k_B$ is the Boltzmann constant. At high enough temperatures, almost all steps are accepted. At lower temperatures, fewer high energy structures are accepted. The simulation proceeds as a series of cycles, each at a specified temperature. Each cycle contains a large number of individual steps, accepting or rejecting the steps based upon the current temperature. After a specified number of acceptances or rejections, the next cycle begins with a temperature lowered by a specified schedule such as:

$$T_i = gT_{i-1}$$

Where $T_i$ is the temperature at cycle $i$, and $g$ is a constant between 0 and 1.

Simulated annealing allows an efficient exploration of the complex configurational space with multiple minima that is typical of a docking problem. The separation of the calculation of the molecular affinity grids from the docking simulation provides modularity to the procedure, allowing the exploration of a range of representations of molecular interactions, from constant dielectrics to finite difference methods and from standard 12-6 potential functions to distributions based on observed binding sites.

3.4 Thermodynamic cycle for enzyme-inhibitor complex

The enzyme and inhibitor complex requires the change in free energy ($\Delta G$) binding under vacuum condition. Enzymes are solvated in water, which needs solvation free energy according to Rommie co-workers (Rommie et al., 2008).
The Fig. 3.2 shows the thermodynamic cycle for the binding of an enzyme, $E$, and an inhibitor, $I$, in both the solvated phase and in vacuum. Note the solvent molecules are indicated by filled circles: they tend to be ordered around the larger molecules, but when $E$ and $I$ bind, several solvent molecules are liberated and become disordered. This is an entropic effect and is the basis of the hydrophobic effect. The solvent ordering around $E$ and $I$, when both bound and unbound, is strongly influenced by the hydrogen bonding between these molecules. These hydrogen bonds between solvent and $E$, and solvent and $I$, contribute enthalpy stabilization.

### 3.5 Hess Law application in docking

According to Hess's law of heat summation, the change in free energy between two states will be the same, no matter what the path. So we can calculate the free energy of binding in solvent by the following equation:

$$\Delta G_{\text{binding solvation}} = \Delta G_{\text{binding vacuum}} + \Delta G_{\text{solvation}(EI)} - \Delta G_{\text{solvation}(E+I)}$$
Since we can calculate $\Delta G_{\text{binding, vacuo}}$ from docking simulation, and can estimate the free energy change upon solvation for the separate molecules $E$ and $I$, and for the complex, $EI$, $\Delta G_{\text{solvation}(EI)}$ and $\Delta G_{\text{solvation}(E+I)}$ respectively, then it is also possible to calculate the free energy change upon binding of the inhibitor to the enzyme in solution, $\Delta G_{\text{binding,solvation}}$. Thus, we can estimate the inhibition constant, $K_i$, for the inhibitor, $I$. A key point to bear in mind is that most parts of the new scoring function are essentially the same as the original AutoDock scoring function used in versions prior to 3.0, except that various terms in the molecular mechanics energy function have been re-scaled by new coefficients, and new terms have been introduced. These new terms include the desolvation free energy of the ligand, and an estimate of the loss of conformational degrees of freedom of the ligand upon binding.

### 3.6 Application of docking to discover new drugs

A closely related challenge is the development of effective methods to predict the binding propensity for series of compounds or flexible peptides to a given receptor (Lybrand et al., 1986; Gilson et al., 1997 and Aquist et al., 2002). Approaches based on scoring functions or compound libraries require large amounts of data to be available a priori. These methods include computational virtual screening (Blake et al., 2000; Egan et al., 2002; Li et al., 2004; Irwin, 2006; Fonovic and Bogyo, 2007), docking (Brooijman and Kuntz, 2003; Taylor et al., 2002; Schulz-Gasch and Stahl, 2003; Cole et al., 2005; Souse et al., 2006; Sperandio et al., 2006) and similarity searching (Whittle et al., 2004; Hajduk and Greer, 2007). As the binding propensity defining a given molecular association reflects the relative stabilities of the possible conformations of the receptor, effective drug-design protocols should be based on a distribution of receptor conformations. Prof. J. Andrew McCammon and his colleagues used AutoDock and the Relaxed Complex Method (RCM) to discover novel modes of inhibition of HIV integrase. Researchers at Merck Pharmaceutical Company have used McCammon's work to design new drugs that target integrase, which lead in October 2007 to the first clinically-approved HIV Integrase inhibitor: Isentress™ (raltegravir). Crucial role of docking in drug design has been proven by expertise in pharmaceutical companies as well as researchers in scientific field of
bioinformatics. The drug molecule conformation in a cavity has shown in Fig 3.3 to calculate binding energy with new bioinformatics and chemoinformatics tools.

![Drug molecule docked into the targeted protein inhibitor site and orientation of drug molecule.](image)

**Fig.3.3.** Drug molecule docked into the targeted protein inhibitor site and orientation of drug molecule.

It is essential to study interactions of newly designed lead molecules with specified receptor or enzymes. Recent work carried out by Anuradha and co-workers have been designed antimicrobial agents for *Mycobacterium tuberculosis* of MurC enzyme. They had studied the new anti tuberculosis agents docked (Autodock 3.0) with MurC enzyme (Anuradha *et al*., 2009). These docking studies are crucial for drug design in less time and cost. So docking studies plays a vital role in computer aided drug design in bringing new effective drugs into the market.

*****